3D and Appearance Modeling from Images

Peter Sturm[1], Amaël Delaunoy[1], Pau Gargallo[2], Emmanuel Prados[1], and Kuk-Jin Yoon[3]

[1] INRIA and Laboratoire Jean Kuntzmann, Grenoble, France
[2] Barcelona Media, Barcelona, Spain
[3] GIST, Gwangju, South Korea

Abstract. This paper gives an overview of works done in our group on 3D and appearance modeling of objects, from images. The backbone of our approach is to use what we consider as the principled optimization criterion for this problem: to maximize photoconsistency between input images and images rendered from the estimated surface geometry and appearance. In initial works, we have derived a general solution for this, showing how to write the gradient for this cost function (a non-trivial undertaking). In subsequent works, we have applied this solution to various scenarios: recovery of textured or uniform Lambertian or non-Lambertian surfaces, under static or varying illumination and with static or varying viewpoint. Our approach can be applied to these different cases, which is possible since it naturally merges cues that are often considered separately: stereo information, shading, silhouettes. This merge naturally happens as a result of the cost function used: when rendering estimated geometry and appearance (given known lighting conditions), the resulting images automatically contain these cues and their comparison with the input images thus implicitly uses these cues simultaneously.

1 Overview

Image-based 3D and appearance modeling is a vast area of investigation in computer vision and related disciplines. A recent survey of multi-view stereo methods is given in [6]. In this invited paper, we provide a brief overview of a set of works done in our group, mainly by showing sample results. Technical details can be found in the relevant cited publications.

3D and appearance modeling from images, like so many estimation problems, is usually formulated, explicitly or implicitly, as a (non-linear) optimization problem[4]. One of the main questions is of course which criterion to optimize. We believe that the natural criterion is to maximize photoconsistency between input images and images rendered from the estimated surface geometry and appearance (to be precise, this criterion corresponds to the likelihood term of a Bayesian problem formulation, which can be combined with suitable priors).

[4] There exist some exceptions in special cases. For example, in basic shape-from-silhouettes, the 3D shape is directly defined by the input and no estimation is necessary, just a computation to explicitly retrieve the shape.

To measure photoconsistency, one may use for example the sum of squared differences of grey levels or the sum of (normalized) cross-correlation scores. This criterion is simple to define but turns out to be hard to optimize rigorously. To optimize it we process a gradient descent. When speaking about gradient descent, a central question is how to compute the gradient of the criterion. Yezzi and Soatto have shown how to do so, but only for convex objects [7]. In [3], we developed the gradient for the general case. Importantly, it correctly takes into account how surface parts become visible or invisible in input cameras, due to the surface evolution driven by the gradient. Hence, using this gradient, silhouettes and apparent contours are implicitly handled correctly since these are the places where such visibility changes take place. Further, due to comparing input with rendered images, color and shading effects are also naturally taken into account. Overall, rigorously optimizing the photoconsistency between input and rendered images, allows to naturally merge stereo, shading, and silhouette cues, within a single framework and without requiring tuning parameters to modulate their relative influence.

This framework was first developed for a continuous problem formulation [3] (we used level sets for the surface parametrization). We then developed it for the case of discrete surface representations, in particular triangular meshes [2] which in practice allow to achieve a higher 3D surface resolution. Also, even when using a continuous setup, in practice the surface representation is finally discretized and the surface evolution requires to repeatedly discretize attributes. It thus seems more natural to directly start with a discrete parametrization and do all derivations based on it. In both cases, continuous and discrete, the surface evolution can be carried out by gradient descent (one may also try less basic methods, such as conjugate gradient, quasi-Newton methods etc.).

The developed framework for optimizing photoconsistency was then used to develop a general purpose algorithm for modeling 3D surface and appearance [8, 9]. Here, we considered the case where lighting conditions are known (we modeled this as a set of point or directional light sources, plus an ambient lighting) but may be different for each input image. The most general instance of our algorithm estimates an object's 3D surface and a spatially varying appearance. For the latter, we use the standard Blinn-Phong reflectance model and can in principle estimate one set of reflectance coefficients (albedo and specular coefficients) per surface point, allowing to reconstruct non-Lambertian surfaces. However, estimating specular coefficient for each point is obviously highly ill-posed, so the most general experiment we carried out used a strong smoothness prior over these coefficients.

This general algorithm can be run on more constrained examples, in principle simply by leaving out the appropriate parts in the problem parametrization and the computation of cost function, gradient, etc. Examples of some scenarios are given in the following section. For example, one may model the surface appearance by a spatially varying albedo plus uniform specular coefficients, by a spatially varying albedo and no specular effects or simply by a uniform albedo. In the case of constant lighting, the second case corresponds to multi-view stereo

whereas the third case corresponds to (multi-view) shape-from-shading. Also, if variable lighting conditions are considered but a static viewpoint, the algorithm will perform photometric stereo, whereas in the general case of varying lighting and viewpoint, one finds a combination of multi-view and photometric stereo.

2 Sample Scenarios and Results

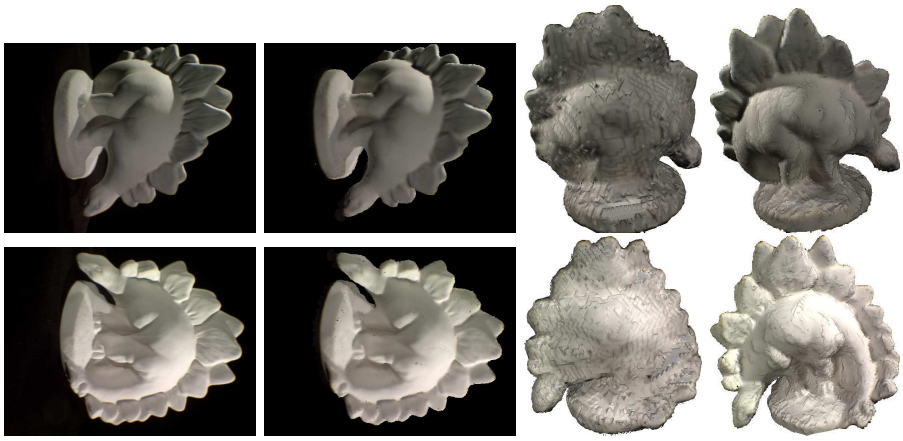
As mentioned above, due to the generality of the proposed approach, it can be applied to various types of image sets with different camera/light configurations. Here, knowledge of illumination allows to factorize radiance into reflectance and geometry. In practice, depending on the scenario, that knowledge may not be required, e.g. for recovering shape and radiance of Lambertian surfaces with static illumination. In other words, when images of Lambertian surfaces are taken under static illumination, the proposed approach can be applied even without lighting information, assuming that there is only an ambient illumination. In this case, the approach works much like the conventional multi-view stereo methods and estimates the shape and radiance of Lambertian surfaces. Figure 1 shows the result for the dino image set [6], for which no lighting information is required. The proposed method successfully recovers the shape as well as the radiance.

In the following, for synthetic data sets, the estimated shape is quantitatively evaluated in terms of accuracy and completeness as in [6]. We used 95% for accuracy and the 1.0mm error for completeness. For easy comprehension, the size of a target object is normalized so that it is smaller than [100mm 100mm 100mm]. Here, beside the shape evaluation, we also evaluated the estimated reflectance in the same manner. For each point on an estimated surface, we found the nearest point on the true surface and compute the distance and reflectance differences, and vice versa.

The proposed approach can also be applied to images taken under varying illumination. Results using images of textureless/textured Lambertian surfaces are shown in Figs. 2 to 5. Figure 2 shows the ground-truth shape of the "bimba" image set (18 images) of a textureless object, and the estimation result. The surface has uniform diffuse reflectance and input images were taken under different illuminations. In this case, the approach works as a multi-view photometric stereo method and recovers the shape and the diffuse reflectance of each surface point. Here, black points in the estimated model correspond to points that were not visible from any camera and/or any light source.

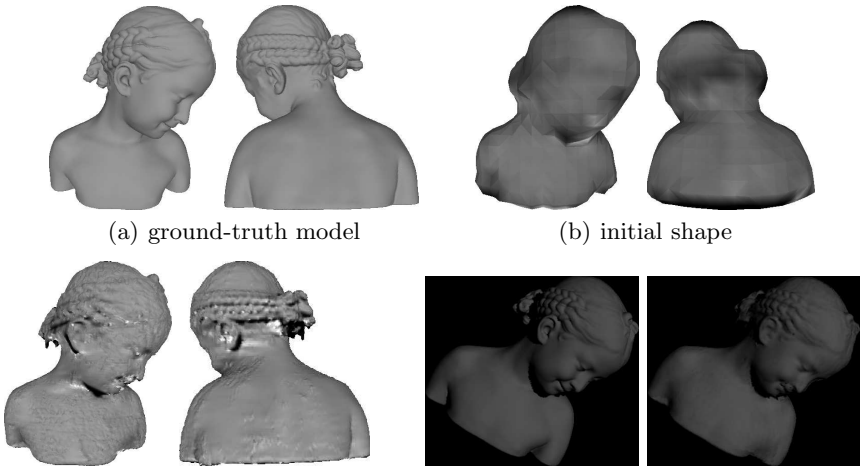
Results for a more complex object are shown in Figs. 3 and 4. The images synthesized using the estimation closely resemble input images while the shading and the reflectance are successfully separated. Furthermore, it is possible to synthesize images under different lighting conditions, even from different viewpoints. The proposed method also recovers concave parts well as shown in Fig. 5.

We also applied our approach to the images of textureless/textured *non-Lambertian* surfaces showing specular reflection. Note that, unlike previous methods [1, 4], we do not use any thresholding to filter out specular highlight pixels.



(a) input images (b) synthesized images (c) initial shape (d) estimated shape

Fig. 1. Result for the "dino" image set (16 images) — Lambertian surface case (static illumination and varying viewpoint).



(a) ground-truth model (b) initial shape

(c) estimated model (d) input image vs. synthesized image

Fig. 2. Result for the "bimba" image set (18 images) — textureless Lambertian surface case (varying illumination and viewpoint). 95% accuracy (shape, ρ_{dr}, ρ_{dg}, ρ_{db})=(2.16mm, 0.093, 0.093, 0.093), 1.0mm completeness (shape, ρ_{dr}, ρ_{dg}, ρ_{db}) = (82.63%, 0.104, 0.104, 0.104).

(a) input image (b) ground-truth reflectance (c) ground-truth shading (d) inital shape

(e) synthesized image (f) estimated reflectance (g) estimated shading

Fig. 3. Result for the "dragon" image set (32 images) — textured Lambertian surface case (static illumination and varying viewpoint). 95% accuracy (shape, ρ_{dr}, ρ_{dg}, ρ_{db})=(1.28mm, 0.090, 0.073, 0.066), 1.0mm completeness (shape, ρ_{dr}, ρ_{dg}, ρ_{db}) = (97.11%, 0.064, 0.056, 0.052).

The result for the smoothed "bimba" data set is shown in Fig. 6. In this case, the surface has uniform diffuse/specular reflectance and each image was taken under a different illumination. Although there is high-frequency noise in the estimated shape, the proposed method estimates the specular reflectance well. Note that most previous methods do not work for image sets taken under varying illumination and, moreover, they have difficulties to deal with specular reflection even if the images are taken under static illumination. For example, Fig. 7 shows a result obtained by the method of [5] and our result for comparison. We ran the original code provided by the authors many times while changing parameters and used mutual information (MI) and cross correlation (CCL) as similarity measures to get the best results under specular reflection. As shown in Fig. 7, the method of [5] fails to get a good shape even when the shape is very simple, while our method estimates it accurately. Also, with such images, given the large proportion of over-bright surface parts, it seems intuitive that the strategy chosen by [1] and [4] (who consider bright pixels as outliers) might return less accurate results, because it removes too much information.

We also used real image sets of textured glossy objects, which were taken by using fixed cameras/light sources, while rotating the objects as in [1, 4] — in this case, each image has a different illumination and observes specular reflections. The light position and color were measured using a white sphere placed in the scene. Figure 8 shows one image among 59 input images, the initial shape obtained using silhouettes, and the final result. Here, we simply assumed a single-material surface (i.e. uniform specular reflectance, but varying albedo). Although a sparse grid volume was used, the proposed method successfully estimated the

(a) ground-truth rendered with a different lighting and from a new viewpoint

(b) ground-truth reflectance

(c) ground-truth shading

(d) reconstructed scene relighted as in (a) and viewed from the same point of view as (a)

(e) estimated reflectance

(f) estimated shading

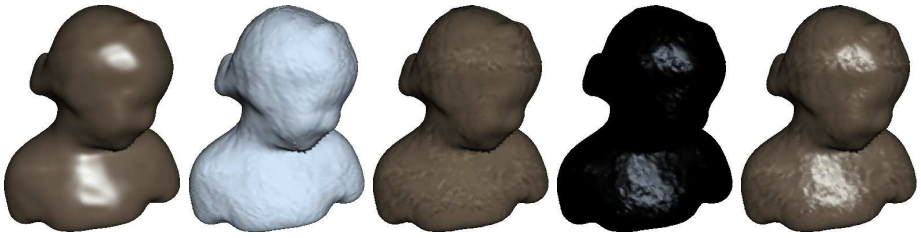
Fig. 4. Synthesized result for different lighting conditions and viewed from a viewpoint that is different from all input viewpoints. A comparison with the ground-truth is possible because this is synthetic data.

(a) ground-truth

(b) close-up view

(c) close-up view of the estimated result

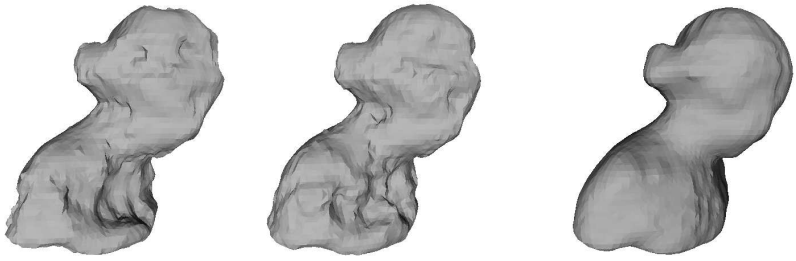
Fig. 5. Close-up view of the concave part of the "dragon" model.



(a) ground-truth model (b) estimated shape (c) diffuse image (d) specular image (e) synthesized image

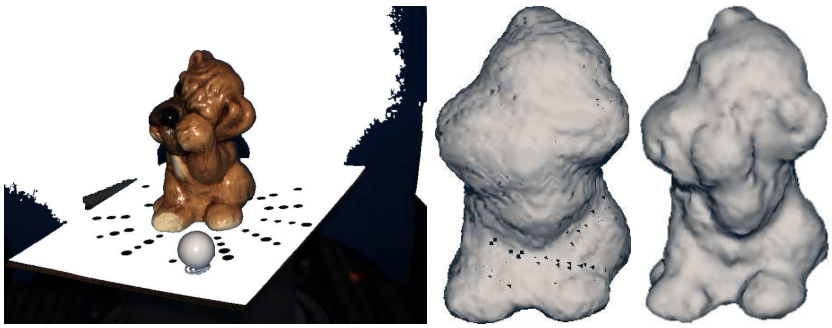
Fig. 6. Result for the smoothed "bimba" image set (36 images) — textureless non-Lambertian surface case (uniform specular reflectance, varying illumination and viewpoint). 95% accuracy (shape, ρ_{dr}, ρ_{dg}, ρ_{db}, ρ_s, α_s)=(0.33mm, 0.047, 0.040, 0.032, 0.095, 8.248), 1.0mm completeness (shape, ρ_{dr}, ρ_{dg}, ρ_{db}, ρ_s, α_s) = (100%, 0.048, 0.041, 0.032, 0.095, 8.248).

(a) two input images

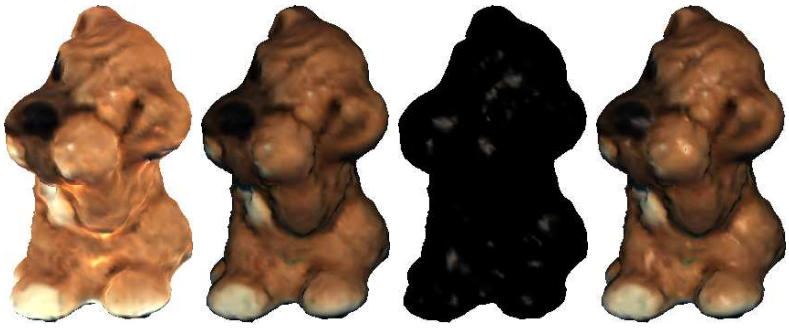


(b) results using [5] (MI and CCL) (c) our result

Fig. 7. Result comparison using the smoothed "bimba" image set (16 images) — textured non-Lambertian surface case (uniform specular reflectance, varying illumination and viewpoint).



(a) input image (b) initial shape (c) estimated shape



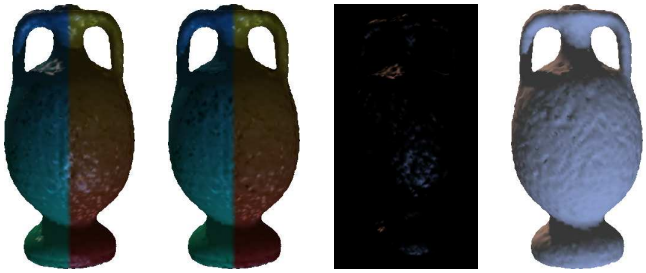
(d) diffuse reflectance (e) diffuse image (f) specular image (g) synthesized image

Fig. 8. Result for the "saddog" image set (59 images) — textured non-Lambertian surface case (uniform specular reflectance, varying illumination and viewpoint).

shape of the glossy object even under specular reflection, while estimating the latter. Here, we can see that, although the estimated specular reflectance may not be highly accurate because of the inaccuracy of lighting calibration, saturation, and unmodeled photometric phenomena such as interreflections that often occur on glossy surfaces, it really helps to recover the shape well.

Finally, we applied our approach to the most general case — images of textured non-Lambertian surfaces with spatially varying diffuse and specular reflectance and shininess, cf. Fig. 9. Input images were generated under static illumination (with multiple light sources) while changing the viewpoint. Figure 9 shows one image among 36 input images, one ground-truth diffuse image, one ground-truth specular image, ground-truth shading, and our results. We can see that the proposed method yields plausible specular/diffuse images and shape. However, there is high-frequency noise in the estimated shape. Moreover, the error in reflectance estimation is rather larger compared to the previous cases because of sparse specular reflection observation. This result shows that, reliably

(a) input image (b) ground-truth diffuse image (c) ground-truth specular image (d) ground-truth shading (e) initial shape



(f) synthesized image (g) synthesized diffuse image (h) synthesized specular image (i) estimated shading

Fig. 9. Result for the "amphora" image set (36 images) — textured non-Lambertian surface case (spatially varying specular reflectance, static illumination, and varying viewpoint). 95% accuracy (shape, ρ_{dr}, ρ_{dg}, ρ_{db}, ρ_s, α_s)=(0.59mm, 0.041, 0.047, 0.042, 0.226, 12.69), 1.0mm completeness (shape, ρ_{dr}, ρ_{dg}, ρ_{db}, ρ_s, α_s) = (89.73%, 0.042, 0.047, 0.042, 0.226, 12.65).

estimating specular reflectance for all surface points is still difficult unless there are enough observation of specular reflections for every surface point.

3 Conclusion

In this paper, we have given a coarse overview of our works on multi-view 3D and appearance modeling. Contrary to previous works that consider specific scenarios, our approach can be applied indiscriminately to a number of classical scenarios — it naturally fuses and exploits several important cues (silhouettes, stereo, and shading) and allows to deal with most of the classical 3D reconstruction scenarios such as stereo vision, (multi-view) photometric stereo, and multi-view shape from shading. In addition, our method can deal with non-Lambertian surfaces showing strong specular reflection, which is difficult even in

some other state of the art methods using complex similarity measures. Technical details are given in our previous publications. Also, although the proposed approach can in principle deal with very general scenarios, especially the case of estimating specular coefficients remains challenging in practice due to numerical issues. A discussion of such practical aspects is provided in [9].

4 Acknowledgements

This work was supported by the Korea Research Foundation Grant funded by the Korean Government(MOEHRD) (KRF-2006-352-D00087) and by the FLAMENCO project (grant ANR-06-MDCA-007).

References

1. Birkbeck, N., Cobzas, D., Sturm, P., Jägersand, M.: Variational shape and reflectance estimation under changing light and viewpoints. European Conference on Computer Vision, I:536–549 (2006)
2. Delaunoy, A., Gargallo, P., Prados, E., Pons, J.-P., Sturm, P.: Minimizing the multi-view stereo reprojection error for triangular surface meshes. British Machine Vision Conference (2008)
3. Gargallo, P., Prados, E., Sturm, P.: Minimizing the reprojection error in surface reconstruction from images. IEEE International Conference on Computer Vision (2007)
4. Hernández Esteban, C., Vogiatzis, G., Cipolla, R.: Multiview photometric stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(3), 548–554 (2008)
5. Pons, J.-P., Keriven, R., Faugeras, O.: Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. International Journal of Computer Vision, 72(2), 179–193 (2007)
6. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. IEEE Conference on Computer Vision and Pattern Recognition, 519–528 (2006)
7. Yezzi, A., Soatto, S.: Stereoscopic segmentation. International Journal of Computer Vision, 53(1), 31–43 (2003)
8. Yoon, K.-J., Prados, E., Sturm, P.: Generic scene recovery using multiple images. International Conference on Scale Space and Variational Methods in Computer Vision (2009)
9. Yoon, K.-J., Prados, E., Sturm, P.: Joint estimation of shape and reflectance using multiple images with known illumination conditions. International Journal of Computer Vision, to appear (2009)